The probability of breaching water quality standards – a probabilistic model of river water nitrate concentrations

Fred Worrall[11], Brandon Kerns[1], Nicholas J.K. Howden[2] and Tim P. Burt[3,4], and Helen P. Jarvie[5],

1. Dept. of Earth Sciences, Science Laboratories, South Road, Durham, DH1 3LE, UK.
2. Dept. of Civil Engineering, University of Bristol, Queens Building, Bristol, UK.
3. Dept. of Geography, Science Laboratories, South Road, Durham, DH1 3LE, UK.
4. School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK
5. Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK.

ABSTRACT

In this study we propose an approach to predicting the probability that river waters will exceed a water quality standard. The study used a two-part generalised linear modelling approach within a Bayesian framework. Binomial regression was used to model the probability that a water quality standard would be exceeded and included two factors - the difference between sampling sites and difference between years of sampling. Using a Bayesian approach meant that information could be drawn from all observations from all sites, across all years, and that all results would come with a measure of uncertainty. Furthermore, although some known factors could not be included in the binomial regression, they could be included using Bayes' rule to enhance and inform the results. This approach was applied to assessing the probability of nitrate concentrations in English river waters exceeding the current nitrate water quality standard of 11.3 mg N /l. The study showed that the Bayesian approach decreased the measures

[1] Corresponding author: Fred.Worrall@durham.ac.uk; tel. no. +44 (0)191 3342995; fax no. +44 (0)191 334 2301

of uncertainty in the predicted outcomes was reduced by an average of 60% and increased the effective sample size by 64%.The best-fit model had a root mean square error (RMSE) of 7.9% which equated to an error of ± 1 sample above the water quality standard for the median site. When interaction of factors could be included, then RMSE decreased to 3.8%. It was not possible to include a diurnal cycle, owing to a paucity of sub-daily sampling, but there was a significant seasonal cycle and so outputs could be adjusted by means of Bayes' rule to predict water quality standard exceedance each month. Comparison with the current method of classification shows no significant difference between five out of the six lowest classifications with only the highest classification being correlated with the estimated exceedence rate. With respect to nitrate in English river waters, the average exceedance rate was 8.3% but was declining at a statistically-significant rate of 0.09 %/yr^2.

1. Introduction

The development of water quality standards has meant that maximum allowable concentrations of water quality determinands have been considered important (e.g., United States' Clean Water Act - USEPA, 1976). An example of this is development of nitrate (NO_3^-) water quality standards in the UK and in the European Union for surface and groundwaters where the maximum allowable concentration was set at 11.3 mg-N/l (50 mg-NO_3/l – EC Directive 1991/676 – Official Journal, 1991). With any of these standards there are a number of issues, as there is not always guidance as to what constitutes "failure". For example, if river-water NO_3^- concentrations were measured once a month for a year and the concentration exceeded 11.3 mg-N/l only once, then would the river be considered to have failed the standard? Alternatively, would it be necessary for the average of the nitrate concentration to be greater than 11.3 mg-N/l? In both cases, either where a number of samples exceed a standard or an average, then further questions immediately arise. Firstly, how many samples are required and

what level of exceedance predicates failure of the NO_3^- standard? For the example, if the site exceeded 1 in 12 samples, what would have happened if only 10 samples had been taken or if the one that exceeded the standard been missed? Secondly, what average is being considered: could this involve an assumption of the distribution, that is, are the data normally or log normally distributed? Even if a non-parametric approach is considered, is the mode or the median to be used? Of course, there is no reason to use a measure of the average, or expected value of the chosen distribution, and indeed a percentile could be used (e.g., 95^{th} percentile). With any of these measures comes the possibility of uncertainty and we would propose that when considering the exceedance of a standard, it is the distribution of the water quality determinand that is being sought, from which appropriate measures (e.g., expected value) should be chosen.

For any water quality monitoring programme, it is always a sample of the water quality that is being taken, yet it is the population of the water quality that we wish to know about. This difference between sample and population means makes inference from our samples to the actual distribution of a water determinand (e.g., nitrate) will involve uncertainty and so there is also a requirement to determine within what uncertainty an answer is acceptable. Within the Water Framework Directive (WFD – EC Directive 2000/60 – Official Journal, 2000), European Union countries are mandated to improve surface waters to both good ecological and good chemical status. The ecological and chemical status is to be monitored in river basin districts with regular sampling and assessment on statistics of that observational data (Carstensen, 2007). However, which statistical measure should be used is not mentioned within the Directive and is left to the choice of individual national regulatory bodies. The different choices of which measure is used makes comparison between member states difficult. McBride and Ellis (2001) and Smith et al. (2001) have proposed a binomial approach in dealing with breaches of the Clean Water Act (USEPA, 1976). The binomial probability distribution is

a discrete probability distribution of the number of successes in a sequence of independent experiments. In this case, each sample collected at a site is considered as being an independent experiment and each of these samplings can be viewed as having one of only two outcomes - success or failure (i.e., whether the measured concentration meets or exceeds the defined standard).

In understanding the distribution of a water quality determinand, the uncertainty in an inference about that distribution will improve with an increased sample size, but time and financial constraints will always limit the number of samples that can be collected at an individual site. Indeed, the improvement of uncertainty decreases with increased sampling and so the return upon increased effort of sampling diminishes, raising the question of sampling efficiency. Simply expanding sampling increases costs disproportionately to value of the added data. This study therefore proposes that a consistent and coherent method is needed by which the most efficient use of the available data is used to provide the maximum information: Bayesian analysis allows more efficient use of public funds in this context. Therefore, we propose that, in understanding whether or not a water quality standard has been breached, a Bayesian model is used.

Qian (2015) has argued that to improve estimation at any one site cross-sectional water quality monitoring data should be used, that is, sites should be compared together in an analysis of variance (ANOVA) – that is a normal distribution of data is assumed and this can be referred to as a general linear model. It is not necessary to restrict ANOVA to a normal distribution, but the method can be generalised to other distributions including a discrete probability distribution (eg. the binomial distribution), and therefore the study's approach was generalised linear modelling. Further, as with general linear modelling within a generalised linear modelling approach, it is possible to consider cross-sectional data in the context of explanatory factors, the interactions of factors, and covariates. Where factors are taken independent categorical

variables and covariates are independent continuous variables. Within the context of water quality sampling, a factor could be the month of sampling, while a covariate could be the river flow at the time of sampling. Furthermore, the binomial modelling can be conducted within a Bayesian framework. Conducting statistical modelling within a Bayesian framework brings several advantages. Firstly, the Bayesian approach can combine sources of information in a coherent and consistent manner and this includes the use of prior information; within the context of water quality studies, the prior information would be from other sites and previous sampling. Secondly, the process is always within a probabilistic framework and results can be presented as probability distributions and so it is easy to give a measure of uncertainty on any prediction. Thirdly, the results of the modelling (the posterior) can be used as the prior information for the next analysis and so, this way, information builds.

. Bayesian approaches nhave been used to support and direct parameter choices and uncertainty analysis in physical water quality models (eg. urban wastewater quality – Dotto et al., 2012). Krueger (2017) used a Bayesian approach to improve interpretation of the low frequency monitoring data typical of national monitoring schemes. Krueger (2017) considered concentrations of dissolved oxygen, orthophosphate and ammonium by multiple approaches using both continuous and discrete probability distributions. However, that approach did not include factors or covariates to improve the inference. Worrall et al. (2019) advanced on this Bayesian approach by considering a Bayesian generalised linear model of specific conductance in surface waters and so this approach combined the elements recommend by Qian (2015) and Krueger (2017). However, by considering specific conductance, Worrall et al. (2019) did not have to consider discrete probability relative to a water quality standard.

Therefore, the purpose this study was to combine the benefits of fitting a binomial distribution to understand the probability of exceeding a defined water quality standard, drawing upon cross-sectional data, its known factors and covariates, and with all the statistical

modelling within a Bayesian framework. The approach developed was applied to understanding water quality in the case of nitrate concentrations of English river waters.

2. Approach and Methodology

2.1. Study dataset

This study considered nitrate data from all English river water sampling sites from 2005 to 2017 that were collected by the Environment Agency (https://environment.data.gov.uk/water-quality/view/landing). The data were selected if they were part of routine monitoring of all river water sites. Data reported as being collected as a response to pollution incidents were not included as these are more likely to include greater values and their additional sampling would distort the sample frequency. The sites include streams, rivers and artificial drainage ditches but not any lakes, ponds or waste effluent discharges. Because of the large number of sample sites, and the variability in sampling frequencies, only sites with at least 10 data points and where those data came from more than one year were included. This selection was, in part, to help the statistical modelling by improving estimation of inter-year variation, but also reduce the number of sites - even after the use of these two criteria there were data from 7942 sites that could be included in the analysis. Two factors were included, firstly, the difference between sampling years, henceforward referred to as the "Year factor", and, secondly, the difference between sampling sties, henceforward referred to as the "Site factor". The interaction between these two terms was considered. The possible inclusion of other factors and covariates is discussed below.

2.2. Binomial regression

The binomial modelling was performed within the well-tested and efficient WinBUGS software (version 14 – Lunn et al. 2013) using Monte Carlo Markov Chain (MCMC) techniques

149 using the Metropolis-Hastings sampler. Prior distributions on the constant and levels of each
150 factor were chosen as vague and as the WinBugs. For the constant term the prior is
151 approximated as normal distribution with mean = 0 and standard deviation = 10, similarly the
152 prior distribution of each level in each factor was set as a normal distribution with mean = 0
153 and standard deviation = 2.5. The length of the MCMC chain was 30,000 cycles after a 10,000
154 burn in cycles with samples saved every 10 cycles and with 1 chain. Model fit was tested using
155 a number of approaches. First, that the 95% credible interval for any factor does not include
156 zero, this is henceforward referred to as being significantly different from zero at a probability
157 of 95%. Note that in a Bayesian analysis the equivalent of the confidence interval in frequentist
158 statistics is the credible interval. Second, that inclusion of the factor, interaction, or covariate
159 caused the total model deviance to decrease – the deviance is a goodness of fit measure and is
160 a generalization of the idea of using the sum of squares of residuals in ordinary least squares.
161 Third, that the inclusion of an additional factor, interaction or covariate decreased the deviance
162 information criterion (DIC). It is generally true that inclusion of factors, interactions or
163 covariates will decrease the total deviance of a model as their inclusion means greater degrees
164 of freedom for fitting, and so the DIC accounts for the inclusion of more fitting parameters
165 against the additional fit of the model by penalising for additional parameters relative to the fit
166 of the model – it is general case of Aikake Information Criterion. Finally, it was also possible
167 to directly compare the fit of the model to the observed data, that is, the predicted exceedence
168 rate a site to the observed exceedence rate at the same site and this comparison was done using
169 the root mean square error (RMSE).

170 No covariates were considered in the initial model development although, for reasons
171 outlined below, Year will later be included as a covariate. Although it is easy to consider a
172 number of plausible covariates, they were not available for each site. For example, most sites
173 were not monitored at river flow gauging stations and so river flow was not available.

2.3. Inclusion of allied factors

The date and time of sampling was known for each observation; however, it was difficult to include these two as factors. Although multiple samples were taken at each site in each year, this was not the case for month of the year or time of day. The typical sampling frequency was only once a month and so there was not multiple sampling within any one month from which a proportion of success and failure of exceedence could be reasonably made. Similarly, with typically only one sample per month per site, the range of the times of day that were included was minimal and so impossible to calculate a sensible distribution of successes and failures for time of day for each site. For these known controls that could not be included in the generalised linear modelling, then Bayes' rule could be applied separately to inform and update inference. By using Site and Year factors in the binomial modelling, the method is capable of predicting the probability of exceeding 11.3 mg N/l at any site included in the analysis for any year in the study period, but by looking across all the observations it was possible to test whether exceedence rates vary between months, and between times of day, and apply these probabilities to modify the predicted probabilities: this can be achieved by straightforward application of Bayes' rule. In its very simplest form Bayes' rule is:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \qquad \text{(i)}$$

where: $P(A|B)$= the probability of A given that B is true; $P(B|A)$ = the probability of B given A; P(A) = the probability of A; and P(B) = the probability of B. In this case $P(A|B)$ is the probability of exceeding 11.3 mg N/l at a site in month B; $P(B|A)$ the probability of it being month B given that 11.3 mg N/l has been exceeded at a site; P(A) the probability of exceeding 11.3 mg N/l at a site; and P(B) is the probability of month B: this latter terms acts to normalise

199 so that $\sum_1^{12} P(A|B) = P(B)$. Likewise, this approach can be viewed in terms of the time of day
200 of sampling and so given a prediction of the probability of exceeding 11.3 mg N/l for a given
201 site in a given year. Then, given knowledge of the seasonal and diurnal cycle in probability of
202 exceeding 11.3 mg N/l, it is possible to estimate the probability of exceeding at any site at any
203 time. The entire dataset was examined to assess $P(B|A)$ where B is month of the year and time
204 of day.
205
206 ***2.4. Validation***
207 As a means of validating and demonstrating the modelling approach several approaches were
208 taken.
209 First, approach was to test the projection of results on the basis of the trend in the data.
210 To do this assessment, the data from 2017 were held back and the model calculated from 2005
211 to 2016 was applied to those new data for that year. To project the model forward to 2017, the
212 Year factor was considered as a covariate instead of as a factor and the binomial model was re-
213 calculated for all data for all sites between 2005 and 2016 using the Site factor but Year was
214 included as a covariate and not as a factor. To assess this projection a binomial model was
215 calculated using only the observations from 2017, and therefore, only Site was used as a factor.
216 The results of the analysis of 2017 were compared to those predicted for 2017 based upon
217 projection from all previous results given year as a covariate.
218 Secondly, to compare to this projection on the basis of the trend using Year as a
219 covariate, predictions for 2016 based upon including site and year as factors (not as a covariate)
220 were compared with predictions of the binomial model based upon results for 2017 alone. This
221 test was used to see whether the probabilities for one year are a good prediction for a subsequent
222 year and performed better than using Year as a covariate, i.e. performing better than basing
223 estimates on the time trend within the data.

Thirdly, to test predictions within the the time series of the available data as opposed to projections beyond the available data, the predictions of binomial modelling for the year 2015 based on using year as factor compared to year as a covariate were compared. This comparison enables to assess the method when year is used a covariate and when used a factor.

As a means of demonstrating the comparison between weak and strong priors the predictions of the model applied only to the data from 2015, were compared to the predictions for 2015 based on all years of data from 2005 to 2016.

2.5. Comparison to current method

With respect to assessment of nitrate concentrations in surface waters in England, the approach used by English regulators has been to use Weibull's method. The Weibull method is a non-parametric approach to estimate the percentiles of a distribution (Ellis et al., 1993). The Weibull method was applied for any site for which there were at least 19 samples over a four-year period (2009 – 2014) leading up to the period assessment with respect to nitrate vulnerable zones. The samples are ordered and 95th percentile is then the concentration in the rth sample where for the 95th percentile this is the value of the rth ranked observation where:

$$r = 0.95(n + 1) \qquad \text{(ii)}$$

where: r = rank of the 95th percentile sample; and n = number of samples. Note that for this approach there must be 19 samples, and that also if the r is not an integer the value of r is rounded down.

It is possible, within this non-parametric Weibull method, to assess the uncertainty in the assessment using an application of the binomial distribution (see supplementary material). The Environment Agency applied the non-parametric Weibull method to English surface running waters and classified sites according to both the 95th percentile and the confidence interval on this estimate – note that they chose to use the 90% confidence interval, (i.e., the interval between the 5th and 95th percentiles). Note that the confidence interval and the credible intervals have distinct definition between frequentist and Bayesian statistical approaches but for practical purposes have the same interpretation. The classification of surface running waters was then based upon the following intervals:

1: at least 95% confidence that the 95th percentile concentration is ≤ 11.3 mg N/l;

2: at least 75% confidence that the 95th percentile concentration is ≤ 11.3 mg N/l;

3: at least 50% confidence that the 95th percentile concentration is ≤ 11.3 mg N/l;

4: at least 50% confidence that the 95th percentile concentration is > 11.3 mg N/l;

5: at least 75% confidence that the 95th percentile concentration is > 11.3 mg N/l;

6: at least 95% confidence that the 95th percentile concentration is > 11.3 mg N/l.

Although these are the rules as set out in Bewes et al. (2015), the actual classification method used in England was not based on nitrate but "total inorganic nitrogen" which was in fact the sum of the nitrate, nitrite and ammonia data available at any site – in this study we applied the classification based on nitrate concentration data alone, as nitrate is the overwhelmingly dominant component of total inorganic nitrogen concentrations in British rivers (Jarvie et al., 2018).

The Weibull method, based as it is on rank position, can only be constrained to give an uncertainty within the bounds of the values measured. In this study, the non-parametric Weibull method was applied to the available nitrate concentration data with respect to 2016 (i.e., based on the years 2012 to 2016) to predict the classification of the sites according to the current

method. For those sites for which the Weibull method was possible, the exceedence rate was also calculated, and the Bayesian binomial modelling approach was applied to predict the probability of exceeding 11.3 mg N/l and the uncertainty in that probability estimate. The output of this study's approach can be compared with the current approach, and therefore the classification and its uncertainty can be judged.

2.6. Examples of Application

The method developed was illustrated by application to three specific sites. The three specific sites were chosen to the cover the spectrum of available information, that is, one site was chosen which only just qualified for this study whereby there was only 10 samples over just two years (River Cole at Box Mill Lane – sampled only in 2005 and 2006). The second site was chosen to present the average site information that is where there were 144 samples over 12 years (River Rother at Canklow). Finally, one of the sites with the largest sampling frequency (River Deben at Cretingham) where there were 472 samples over 12 years. Within these categories, sites were chosen where they were close to average exceedence; after all there would little point demonstrating this method for sites where there was either never any observations above 11.3 mg N/l or for sites where the observations were predominantly above 11.3 mg N/l.

3. Results

In total there 921117 measurements of nitrate at 11476 sites in routine monitoring in England. When the constraint of the site having to be sampled in more than one year is applied there were 639,508 measurements of nitrate in the routine, planned monitoring of English running surface waters for 7,942 sites which qualified under the criteria set with 95,352 combinations of site and year (Figure 1). Of the 639,508 nitrate measurements, 53,258 concentrations were at, or exceeded, 11.3 mg N/l, that is an overall probability of exceeding 11.3 mg N/l of 0.083

298 (8.3%) – henceforward, the probability of exceeding the defined water quality standard will be
299 referred to as the "exceedance rate". Given that the median and modal number of samples per
300 year was 12, then an observed exceedence rate of 8.3 % is equivalent to one exceedence per
301 year per site. The observed exceedence rate per year, regardless of the site, varied from 6.4%
302 in 2008 to 9.9% in 2011.
303
304 ***3.1. Binomial modelling***
305 Fitting a binomial model to the data with only the Year factor showed that there were significant
306 differences (judged by the 95% credible interval) between years. However, when Year was
307 used as a covariate, then there was a significant decrease in the exceedence rate over the course
308 of the period of the study period with the average rate of decrease being 0.09% /yr^2, that is,
309 declining from a mean annual exceedence rate of 8.8% in 2005 to 7.8% in 2016. The value of
310 the Deviance rises from 131 when Year is included as factor to a Deviance of 1,006 when Year
311 is included as a covariate. The fit of the model with Year as a factor gave an RMSE of 14.6%;
312 note this is the percentage error on exceedence rate as a percentage and so at the average
313 observed exceedence rate of 8.3% then this RMSE is ±1.2%. Therefore, at the average site, the
314 fit of the binomial model, using only the Year factor, is equivalent to ±1 exceedence per year
315 per site. The annual predicted exceedence rate could also be expressed as varying from a
316 minimum of 6.4 ± 0.3% in 2008 to a maximum of 10.0 ± 0.3% in 2011 - where the uncertainty
317 is expressed as the 95% credible interval on the annual exceedence rate. However, changes
318 over time in this dataset may be due to changes in the pattern of sampling between sites and
319 times of year. For example, over the period since 2005, sampling moved to focus on sites and
320 times of year where detecting concentrations in excess of 11.3 mg N/l were more likely and so
321 actually mask a larger decline in exceedence rate. Although it was possible that inclusion of
322 individual sites changed over the period of the study, the mean length of record for any site

was 7.5 years and there was no significant change in the overall number of sites that were sampled in each year – the minimum number was 3,645 in 2013 with a maximum of 6,415 in 2006. Although there is no direct evidence for selective sampling inclusion of site as a factor in this analysis would therefore improve the analysis of the trend over the period of the study. However, first we will consider the Site factor alone.

When only the Site factor was considered, the mean annual exceedence rate varied from 0.008 to 98.7%. The RMSE when considering only the Site factor was 8.2%, which for the average site would be an error of ±0.7 exceedences per year.

Including both the Site and Year factors did lead to a reduction in the Deviance and a reduction in DIC; however, the RMSE did not change, remaining at 8.2%, and so we conclude that the majority of the variance in the model is explained by the Site factor and only as small amount is explained by the variation between years. However, the variation between years was now independent of site variation. When the Site factor was then considered, it was independent of the sampling differences between years and the predicted range was between 0.006% and 99% - it should be noted that the binomial model will not predict either 0% or 100%. The distribution of the predicted exceedence rate when compared with the observed exceedence rate shows a particular problem with this type of data (Figure 2). The observed data are discrete in that it is only possible to either observe or not observe a sample over 11.3 mg N/l and it is not possible to measure half a exceedence. Given that the median and modal number of samples per site per year was 12, then 1 exceedence per year is 9%. Correspondingly, two exceedences per year is 17%, and the distribution of observed results (Figure 2) shows that indeed peaks in the distribution are focused on 8 and 16%. However, the binomial model predicts a continuous probability and gives an expected value of 8.2%. This contrast between the scales upon which data can be observed and data can be calculated

represents a fundamental problem and almost a fundamental limit to the fit of any probability model, if fit is defined in terms of RMSE.

The next advance was to include the interaction term between Site and Year factors; this interaction term represents the nature of the trend at individual sites. However, due to the nature of the sampling it was not possible to estimate the interaction term at each site. Of the 95,352 possible site-year combinations since 2005, there were 13,188 site-year combinations where it was possible to estimate the interaction term. For the same dataset of sites and years of data, when the interaction term was included, the Deviance fell to 4,577. For these same sites and years, the Deviance was 6,924 (for model using Site factor only) and 5,488 (for model using Site and Year as factors but not their interaction). The RMSE of the interaction of 3.2%, that is, at the median sampling rate this approach would be out by one exceedence every 3 years. If the best combination of models was used, i.e. accepting estimates for those where an interaction term could be used wherever possible but otherwise the estimate based upon the Site and Year factors only was taken, and in which case the RMSE decreased to 7.9%, that is, not much impact over using only factors and not their interaction. Alternatively, directly comparing the exceedence rate estimated from Site and Year factors with the exceedence rate estimated from Site and Year factors including their interaction give a significant(probability of intercept and gradient being different from zero at probability of at least 95%) equation:

$$\theta_{SY} = 0.01 + 0.93\theta_{S+Y} \qquad n = 13187,\ r^2 = 0.89 \qquad \text{(iii)}$$

$$(0.0007)\ (0.002)$$

where: θ_{S+Y} = exceedence rate estimated from Site and Year factors; and θ_{SY} = exceedence rate estimated from Site and Year factors including their interaction term. The values in the brackets below Equation (iii) are the standard errors in those terms and thus it is easy to see that the

gradient of Equation (iii) is significantly different from the unity and so the fit of this equation implies that on average the $\theta_{S+Y}$is an overestimate of θ_{SY}.

It is possible to map the predicted probability and uncertainty on those predictions for any year in the dataset. For the year 2015 (Figure 3) the mean probability varied from 94 to 0.1% with credible intervals varying from 92% (±46%) to 0.2% (±0.1%), that is, for some sites there almost complete uncertainty as to the probability of exceeding 11.3 mg N/l while for others there is almost complete certainty.

The site that was the most certain is the River Meon upstream of the Warnford fish farm where 133 samples over 11 years have an average of 6.49 mg N/l and a coefficient of variation of 8.0%. The most uncertain site was Lunt Meadows where 10 samples were taken over 3 years and had an average of 10.1 mg N/l and a coefficient of variation of 40.0%. The comparison between the credible interval on the mean probability of exceeding 11.3 mg N/l and the sample size shows a significant relationship (Figure 3) with the credible interval declining in proportion to $1/\sqrt{n}$. However, two groups are apparent in the data, the two groups are:

$$CI = \frac{0.34}{\sqrt{n}} \qquad R^2 = 0.86 \qquad \text{(iv)}$$

$$CI = \frac{1.26}{\sqrt{n}} \qquad R^2 = 0.59 \qquad \text{(v)}$$

where: CI = 95% credible interval. When each site was assigned to either Equation (iv) and (v) and plotted a spatial pattern emerges (Figure 4) whereby those following Equation (iv) are in areas dominated by permeable geology while those sites in areas dominated by impermeable geology follow Equation (v). For those sites following Equation (iv) the improvement (i.e. decrease in CI), is smaller with the same increase in sample size when compared to those sites

following Equation (v), for example an increase from a sample size of 6 to 12 would mean a 29% decrease in the CI for both those following Equation (iv) and (v), but for those sites following Equation (v) this an actual decrease of this would be from ±0.26 to ±0.16 while for Equation (iv) it would ±0.07 to ±0.05. However, the CI is always lower for sites following Equation (iv) than those following Equation (v) except for the very highest values of n, i.e., CI is lower for sites dominated by permeable geology and groundwater flow and we can interpret this as being due to the buffering effect of groundwater flow paths making inherently less variable than monitoring sites on impermeable geology.

By comparing results for 2015 based on a strong or weak prior information, the results of the strong prior, i.e., results for 2015 based on all years of available data, were on average 17% higher than those of the weak prior (Figure 5a). Comparing the uncertainty on the estimates then standard deviation based on the strong prior is on average 60% smaller than that based upon the weak prior (Figure 5b), i.e., the more information included in the analysis the more precise the result. However, it should be noted that, in some cases, the precision of the estimation based upon the strong prior was worse than that for the weak prior. Expansion of the distribution around the estimate will occur when there is disagreement between the prior and the observed information which in this case means that the data for 2015 was different from that in the preceding years. Given Equations (iv) and (v) a 60% improvement in the magnitude of the standard is equivalent to a 64% increase in sample size, i.e., for the average site with 12 samples in 2015 the uncertainty in when using strong prior information is equivalent to having 19 samples in that year.

3.2. Comparison to current method

A comparison of the classification based upon the Weibull approach and the exceedence rate at each site (based upon only data over the same time period as the Weibull approach) was used

421 to classify the data. This showed that, although the exceedence rate increased with the
422 classification, there was in fact no significant difference between classification 1 through 5 and
423 only those sites with classification of 6 were actually significantly more likely to have detect
424 above 11.3 mg N/l (Figure 6).
425
426 ***3.3. Inclusion of allied factors***
427 When the frequency of exceedence was considered across the months of the year, then there
428 was a significant seasonal cycle in exceedence rate across the year (Figure 7). The highest
429 exceedence rate was found for November (9.5 ±0.2%) but this was not significantly different
430 from the exceedence rate in January but was significantly different from the month with the
431 third highest exceedence rate (October – 9.0 ± 0.2%). The months of February, March, May,
432 September, October, and December were not significantly different from each other, but this
433 group of months were significantly different from the months showing the lowest rates of
434 exceedence (April, June, July, and August) with the lowest rate of exceedence in July (7.2±
435 0.2%). In contrast to the significant seasonal cycle there was no significant diurnal cycle
436 discernible within this dataset – high-frequency studies of individual rivers have demonstrated
437 diurnal cycles in the nitrate concentration (Halliday et al., 2013). The frequency of sampling
438 varies considerably across the diurnal cycle with only a few hundred samples per hour taken
439 after 18:00 hours and 08:00 hours, i.e., the only sampling that occurs overnight is from those
440 sites where an automatic sampler has been deployed and there were just four such sites across
441 England during the study period – only 1% of the sampling that could be considered by this
442 study were taken between 18:00 and 0800 hours. Between the 0800 and 1700 hours there are
443 several thousand samples an hour peaking for between 1100 and 1200 hours when 20% of all
444 samples considered in the study were taken.
445

3.4. Validation

Exceedence rates were compared for 2017 between data from only that year and, as predicted, based upon the time trend across all the datasets from 2005 to 2016. The exceedence rate based upon observations from only 2017 was on average 33% higher than that predicted on the basis of the trend. When estimates for 2016 were compared to those observed for 2017 then the estimates for 2016 were 12% higher than observed for 2017. When the comparison was made between results calculated for 2015 based on the Year as a factor and Year as a covariate shows that estimates based upon Year as a covariate were on average 20% higher than those when Year was used as a factor. This comparison suggests that, where no particular pattern of behaviour was observed, then the previous year would appear to be a better predictor of the current year than using the trend in the data.

3.5. Application

When applied at the low sampling frequency site, the method was able to predict the exceedence rate across the entire study period although, in comparison to the predicted probabilities of exceedence at the other sites, the uncertainty on the median estimate is very large (±20% - Figure 8a). For the sites with higher total sampling, the exceedence rate follows the pattern of annual and seasonal cycles in the nitrate observations, with median exceedence rate increasing at times when observed concentrations increased (Figure 8b). It should be noted that a direct comparison between the observed concentrations and the predicted probabilities is not possible. For the highly sampled site (Figure 8c) it was reasonable to compare the observed annual exceedence rate to the predicted annual exceedence rate (Figure 9) and in which case we might conclude that the years 2007 and 2011 were unusual.

4. Discussion

This study has proposed that a binomial model approach can work well with current water quality regulation and monitoring schemes. An alternative would be the use of a negative binomial distribution whereby observations are continued until a specified number of failures have been observed, In the case of nitrate concentrations this could mean that the probability of getting one sample above the 11.3 mg N/l standard in a fixed number of observations could be predicted. Such as approach would require each sampling location to be visited a pre-arranged number of times. Rather than predicting the probability of a sample being over the limit or the number of samples that will be over a defined limit the actual distribution of the nitrate concentration and base assessment on that distribution. The current approach used by the Environment Agency in England is in effect a reconstruction of the distribution without stating what that distribution would be – a non-parametric approach. Thus, an alternative here would not be to use Bayesian analysis to estimate the probability of exceeding 11.3 mg N- /l but rather to estimate the distribution of the nitrate concentration at each monitoring site based upon the observed data but by using a Bayesian approach drawing upon all observations across all monitoring sites to improve the inference at any one site. Once the distribution at each site has been estimated then any number of metrics could be read from it and this would not be restricted to non-parametric measures such as median. However, such an approach does introduce a second-order uncertainty in to the problem, that is, what is the correct distribution (e.g., a gamma or normal distribution). Within Bayesian generalised linear model development assessing which of a range of distributions is applicable is relatively easy to assess as the model deviances can be compared. Krueger (2017) did try different disstributions, both discrete and continuous, in their study of a range of water quality determinds. However, is one distribution type equally applicable to all sites across an area the size of England, for example, would a log-normal distribution be a best-fit everywhere or would not a mixture of gamma and log-normal be better depending upon the characteristics of the site?

Could we have improved on the method presented here? The estimation presented here would improve with additional factors, their interactions or covariates. This study has already considered, but could not use, hour of the day as a factor but in interaction with the Site factor it might be possible to include the estimate of the diurnal cycle at each monitoring site. The interaction between the Site and Year factors has been shown to greatly improve the uncertainty in the prediction of the probability and so any improvement in sampling that ensure this interaction term can be estimated would greatly improve this model. In this study so far, no covariates have been included but the method could readily accommodate covariates if suitable covariates could be found. One obvious covariate for assessing probabilities in river water quality would be the river discharge. It is easy to develop a hypothesis based upon the relationship between nitrate concentrations and river discharge, for example, where concentrations decline or increase with higher river discharges, river discharge should be a good covariate; however, only 675 of the monitoring sites included here were actually co-located with river gauging stations. Figure 4 also suggests the nature of the underlying aquifer in the monitored catchments maybe a useful categorical variable. Other catchment properties may also be useful, e.g., rainfall, but all catchment properties are precisely collinear with monitoring site. So if Site is included as factor in any statistical model then catchment properties are irrelevant, however, if the aim were to extend predictions to unmonitored catchments then relationship to catchment properties will be useful. More covariates may come from other water quality parameters measured at the same monitoring sites as included for nitrate in this study. Other water-quality determinands that might be expected to covary because of correlated biogeochemistry, or parameters could be chosen because they share hydrological pathways. To exploit the availability of covariates and the possibility of estimating the interaction of factors, an alternative strategy would be to focus upon a subset of sites at river gauging stations.

521 This method was applied to the nitrate concentration of English river waters and showed
522 that over the period of the study (2005-2017) there was a significant decline in the rate at which
523 the water standard was being exceeded, that is, relative to the water quality standard English
524 river waters had improved. The UK has a number of measures to help mitigate and reduce
525 nitrate pollution of surface and ground waters and Nitrate Pollution Prevention Regulations
526 were brought in in 2008 and updated in 2015 (Nitrates Pollution Prevention Regulations, 2015)
527 to help implement the EU Nitrates Directive (Official Journal of the European Commission,
528 1991) by extending the area of the country that was designated as a nitrate vulnerable zone
529 (NVZ). As of 2017, 58% of the England's land area (76000 km^2) is now within an NVZ
530 (Environment Agency, 2016). This study has not used designation of the sites as a covariate or
531 conversely used the exceedence rates to assess the effectiveness of site designation in
532 controlling nitrate pollution of river waters. However, the comparison between this study's
533 binomial modelling approach and current, non-parametric Weibull method shows that the latter
534 was overly sensitive below the highest classification, due to a method that was insensitive at
535 the relatively small numbers of samples that could be considered at most sites. However, a
536 better comparison would be to compare estimates from the modelled distribution of the nitrate
537 rather than probability of exceedance. Modelling to nitrate concentration could include the
538 same set of factors and potential covariates as used in this study for modelling the probability
539 of exceedance and use a range of distributions that suit nitrate concentration data distributions
540 which are both truncated at 0 mg N/l and by analytical detection limit.
541

542 **6. Conclusions**

543 This study developed a two-part Bayesian binomial regression model for estimating the
544 exceedence rate of exceeding a given water quality standard. The approach draws upon cross-
545 sectional data allowing for observations at different sites and across differing years to improve

546 inference about the rate of exceeding a given water quality at any one site. This study has shown
547 the advantage of a Bayesian approach: by drawing upon information from multiple years and
548 multiple sites, this improves precision of exceedence rates. Even when factors could not be
549 included in the binomial regression, they could be considered via a two-part application of
550 Bayes' rule. In application to the nitrate concentrations of English river waters, the analysis
551 showed a significant decline in detecting nitrate above the agreed water quality standard. The
552 use of statistical models within a Bayesian framework provides a coherent and consistent
553 method for handling monitoring data making maximum use of public funds.
554

555 **Acknowledgements**

556 The data for this study were all taken from the Environment Agency's WIMS database and all
557 of which are available freely from the UK government.
558

559 **References**

560 Carstensen, J. (2007). Statistical principles for ecological status classification of Water
561 Framework Directive monitoring data. Marine Pollution Bulletin 55, 3-15.
562 Dotto et al. (2012) Comparison of different uncertainty techniques in urban stormwater
563 quantity and quality modelling. Water Research 46, 8, 2545-2558.
564 Ellis, J.C., van Dijk, P.A.H. and Kinley, R.D. 1993. Codes of Practice for Data Handling –
565 Version 1. WRc Report No. NR 3203/1/4224 for The National Rivers Authority.
566 Environment Agency, 2016. Review of Nitrate Vulnerable Zone designations for
567 implementation in 2017. Environment Agency, Bristo, England, Report Number
568 HOEV151604/R.

569 Halliday, S.J., Skeffington, R.A., Wade, A.J., Neal, C., Reynolds, B., Norris, D., Kirchner,
570 J.W., 2013. Upland streamwater nitrate dynamics across decadal to sub-daily timescales: a
571 case study of Plynlimon, Wales. Biogeosciences 10, 12, 8013-8038.
572 Jarvie, H.P., Smith, D.R., Norton, L.R., Edwards, F., Bowes, M.J., King, S.M., Scarlett P.,
573 Davies, S., Dils, R., Bachiller-Jareno, N. (2018). Phosphorus and Nitrogen Limitation and
574 Impairment of Headwater Streams Relative to Rivers in Great Britain: A National
575 Perspective on Eutrophication. Science of the Total Environment, 621, 849-862.
576 Krueger, T. (2017). Bayesian inference of uncertainty in freshwater quality caused by low-
577 resolution monitoring. Water Research 115, 138-148.
578 Lunn, D., Jackson, C., Best, N., Thomas, A., Speigelhalter, D., 2013. The BUGS book – A
579 practical introduction to Bayesian analysis. CRC Press, Abingdon, UK.
580 McBride, G.B., Ellis, J.C., 2001. Confidence of compliance: a Bayesian approach for percentile
581 standards. Water Research 35, 5, 1117-1124.
582 Nitrate Pollution Prevention Regulations, 2015. Depaartment of Environment Food and Rural
583 Affairs (DEFRA), England, http://www.legislation.gov.uk/uksi/2015/668/regulation/1
584 (accessed January 2019).
585 Official Journal of the European Commission, 1991. Concerning the protection of waters
586 against pollution caused by nitrates from agricultural sources. Directive 1991/676/EC.
587 Official Journal of the European Commission, 2000. Establishing a framework for Community
588 action in the field of water policy. Directive 2000/60/EC
589 Qian, S.S., 2015. The frequency component of water quality criterion compliance assessment
590 should be data driven. Environmental Management 56, 24-33.
591 Smith, E.P., Ye, K., Hughes, C., Shabman, L. 2001. Statistical assessment of violations of water
592 quality standards under section 303(d) of the clean water act. Environ. Sci. Tech. 35, 1, 606-
593 612.

USEPA, 1976. Quality criteria for water. United States Environmental Protection Agency, Washington DC, USA.

Worrall, F., Wade, A.J., Davies, R.J., Hart, A. (2019). Setting the baseline for shale gas – Establishing effective sentinels for water quality impacts of unconventional hydrocarbon development. Journal of Hydrology 571, 516-527.

603 Figure 1. Location of sites for which sufficient data were available for this study.

604

605 Figure 2. The distribution of the predicted mean probability of exceedence in comparison to
606 the observed exceedence rate (total n for this study was 53,258)

607

608 Figure 3. a) Map of the mean probability of detecting nitrate above a concentration of 11.3
609 mg N/l (the exceedence rate) for 2015. The credible interval of the predictions of exceedence
610 in 2015 with b) the 2.5% lower credible limit; and c) the 97.5% upper credible interval.

611

612 Figure 4. Spatial distribution of the sites distributed by those following Equation (iv) or (v).

613

614 Figure 5. Comparison of a) mean probability of a sample exceeding 11.3 mg $N\text{-}NO_3$/l, and b)
615 the 95% credible interval, estimated based on data only from 2015 (weak prior) or based
616 upon all data (strong prior).

617

618 Figure 6. Comparison of classification of sites by the Weibull approach and the predicted
619 mean probability of a sample exceeding 11.3 mg N/l.

620

621 Figure 7. The annual cycle by month in the probability of a sample exceeding 11.3 mg N/l.

622

623 Figure 8. Application of the modelling approach to: a) River Colne at Box Mill Lane – low
624 frequency sampled site; b) River Rother at Canklow – the average frequency sampled site;
625 and c) River Deben at Cretingham – the high frequency sampled site.

626

627 Figure 9. Comparison of the annual observed exceedence rate for the River Deben at
628 Cretingham and the predicted exceedence rates from the binomial model.